Variable Window Gabor Filters and Their Use in Focus and Correspondence

Yalin Xiong Steven A. Shafer

CMU-RI-TR-94-06

The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

March, 1994

A short version of this technical report will appear in the *Proceedings of Computer Vision and Pattern Recognition 1994*.

Contents

List of Figures

List of Tables

Abstract

More and more low level vision algorithms are being carried out in the spatial frequency domain, using Gabor filters. There are two basic problems concerned with Gabor filterings we will address in this paper. One is the window size problem, in which we will adopt a set of 2D variable window Gabor filters, and compare its performance with those of fixed window filters. We will show that the variable window scheme is more adaptive to image contents, while fixed window schemes may suffer either large errors or instabilities when image contents are changed. The other problem we will address is the stability of amplitude and phase information resulting from convolving the filters with images. We will extend Fleet's 1D phase stability analysis to 2D phase and amplitude stability analysis based upon the assumption of local resemblance of filter outputs to a single sinusoid. Applications on focus quality measurement and 2D correspondence are described, and the results demonstrate improvements of performance by detecting unstable information using the criterion developed.

Keywords: Computer vision, Low-level processing, Gabor filter, Depth from defocus, Depth from stereo, 2D correspondence.

1 Introduction

This paper addresses two problems raising in using Gabor filters to extract information from images. The first problem is the window size problem, in which we use a set of variable window Gabor filters to extract information at each frequency band. The benefit of this set of filters is that we no longer need to adjust the window size parameter for every experiment. The other problem we will address is about the stability problem caused by finite windows. We will extend Fleet's 1D phase stability analysis to 2D phase and amplitude stability analysis, and develop a stability constraint in a more general way. Finally, we will demonstrate the use of the tools in solving two important vision problems, focus and correspondence.

In traditional approaches, the window size for low level operators is tuned manually for specific experiments. Such a tuning process is usually undesirable for an algorithm to be flexible and stable. Intuitively, to extract information of low frequency, large windows are essential for the information to be stable, while for higher frequency components, smaller windows are preferred to preserve locality. This suggests a variable window scheme which decomposes a fixed window into a set of windows whose sizes are directly related to their frequencies. A set of self-similar, rotation and translation invariant 2D Gabor filters, which cover the whole spatial frequency plane up to the Nyquist frequency, has been used to extract information from images. By combining results from different frequency bands, a focus algorithm based on such decomposition demonstrates greater flexibility and adaptiveness than those based on the fixed window size scheme.

It is well known that windowing in the spatial domain is equivalent to convolving in the frequency domain, which limits overall frequency resolution. When amplitude and phase information are extracted from windowed filtering, it is possible that they are severly contaminated by the convolution. Therefore, if any algorithm makes use of that information implicitly or explicitly without any further examination of the stability, it is subject to either large error or total failure. One goal of this paper is to provide a set of constraints which are capable of identifying the contaminated information, and a generic framework for using information.

Focus quality measurement is a typical problem using Fourier amplitude information. The problem can be formulated as measuring the change of amplitude between two images. Applying the set of 2D variable window Gabor filters and the technique of identifying unstable amplitude information, we can show that the performance of the algorithm can be improved in stability, adaptiveness and precision.

2D matching is another typical problem which can be formulated using phase information. Because shifting in the spatial domain is equivalent to phase change in the frequency domain, we can measure phase changes to infer the spatial shift between two images. As the same as we do in focus quality measurement, we will apply the set of filters and the technique of stability analysis to the problem, and the performance of the algorithm can be improved.

It is worth noting that the generic framework of using information extracted from

the set of filters is not limited to any specific application. While in this paper, we demonstrate its application on focus quality measurement and 2D intensity image matching, the framework can certainly be extended to any vision algorithm which makes use of either amplitude or phase information.

2 Related Research

As an alternative way of performing visual computing in spatial domain, the spatial frequency approach has been favored by many researchers for the applicability of various signal processing techniques [12] and biological evidence [4] Previous research on visual computing in the frequency domain has been concentrating on four areas, namely, motion analysis, stereo matching, texture analysis, and focus measure.

Adelson and Bergen [1] and Heeger [9] modeled motions in 2D image space as orientations in 3D spatiotemporal space, therefore, introduced 3D oriented filters to measure image velocities. More recently, Fleet and Jepson [6] modeled the normal velocity as a function of local phase changes, and then use Gabor filters to measure changes of phase at every pixel location. To the stereo matching problem, Weng [20], Sanger [18], Fleet et al. [7], and Langley et al. [15] proposed to use filters to extract phase information, then compute disparities from them, while Jones and Malik [11] applied a set of linear spatial filters to images, and use responses from those filters as matching features. The spatial frequency approach also achieved great success in texture segmentation [3, 10, 14], and shape recovery from texture [13, 16].

One of the major disadvantages of the spatial frequency approach is the artifact introduced by windowing, which may cause substantial error if unnoticed. The usual way to overcome this problem has been to use large windows so that the artifact is negligible, but in price of severely reduced resolution. Fleet and Jepson [5] provided an excellent way to analyze the stability of phase information. One of the major goals of this paper is to generalize their work to stability analysis of both phase and amplitude information.

Measuring focus quality through spatial frequency analysis has been proposed in the literature [22, 2, 19, 17]. Few of the reported results have addressed the stability problem of amplitude information, which is the only information used in focus quality measurement. We will show that the stability analysis is capable of eliminating unstable amplitude information, therefore improving the performance of focus quality measurement.

3 Variable Window Gabor Filters

Limited by the uncertainty principle [8], any filter must compromise between spatial resolution Δx and spatial frequency resolution Δf [1]. The Gabor filter, which is a

[1] In this paper, f always refers to an angular frequency, i.e. the wavelength $\lambda = \frac{2\pi}{f}$

complex sinusoid modulated by a Gaussian function, is the one which can achieve minimal product of spatial uncertainty and frequency uncertainty:

$$G(x, f; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} e^{-jfx}, \tag{1}$$

where the spatial extent σ decides the tradeoff between the spatial resolution and the frequency resolution.

If the spatial extent σ is constant across different frequency bands, i.e. we compute the spectrogram[22], the filters for low frequency bands will have much smaller numbers of waves than those for high frequency bands. Therefore, the spatial localization will be effectively reduced at high frequency, while in spectral domain, the spectral localization in term of octave (logarithmic frequency band) will be effectively reduced at low frequency. Suppose the spatial extent σ vary linearly with the wavelength of the tuned frequency, i.e.,

$$\sigma = \frac{k_1}{f}, \tag{2}$$

we then obtain a set of variable window Gabor filters:

$$G(x, f) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2 f^2}{2k_1^2}} e^{-jfx} \tag{3}$$

Extending the 1D variable window Gabor filters to 2D, we have the 2D Gabor filter:

$$G(x, y, u, v) = g(x, y) e^{j(ux+vy)}, \tag{4}$$

where $g(x, y)$ is an elliptical 2D Gaussian function in general, and (u, v) is the 2D peak frequency of the filter.

The radial frequency of this filter is $f = \sqrt{u^2 + v^2}$ and orientation $\theta = \tan(v/u)$. Also it is usually more convenient to have the modulating elliptical Gaussian with the same orientation θ as the filter. Figure 1 illustrates the peak frequency position, the spectral extent of one filter in the frequency domain, and the real part of the filter in the spatial domain. A remaining free parameter is the aspect ratio of the elliptical Gaussian. Inspired by some biological evidence reported in [4], we choose the aspect ratio as two-thirds.

As we did for the 1D Gabor filter, we constrain the spatial extent in the θ direction of the 2D filter to be proportional to the radial wavelength, we then obtain a set of 2D Gabor filters which are translated, rotated, and dilated or contracted versions of each other. In the experiments we show below, we used a set of 120 2D filters, which have 10 different radial frequencies and 12 orientations. The k_1 in Eq. 3 is set to π. The spectral extents of filters are described in Figure 2, assuming the extent of $g(x; \sigma)$ is from $-\sigma$ to σ.

As we will show in experiments, the benefit of using such a set of filters with different size of spatial support is that when results from different frequency bands are combined, if the image has strong high frequency components, the final result will be strongly influenced by results from high frequency bands, and therefore, it is

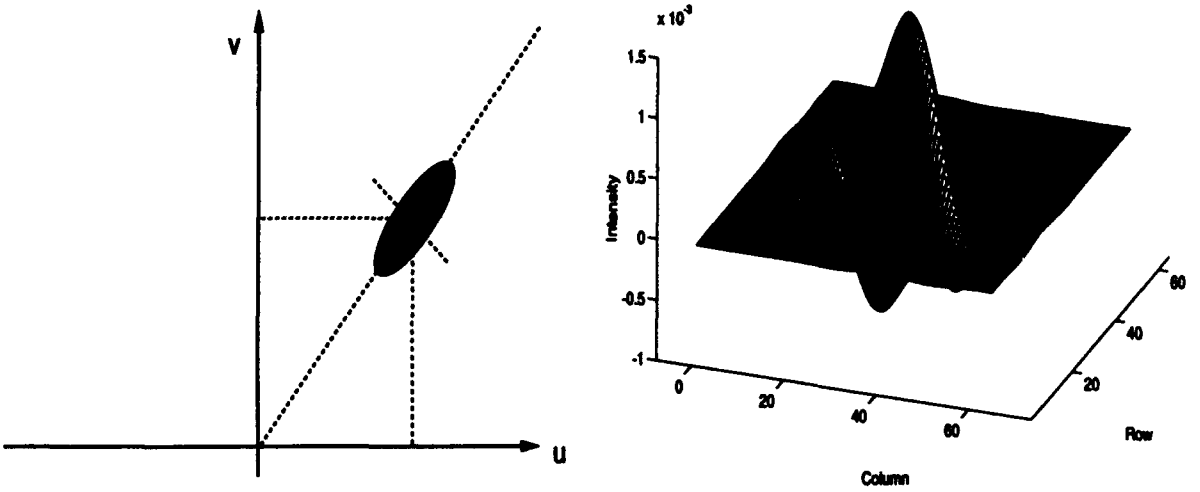


Figure 1: Oriented 2D Gabor Filter in Frequency and Spatial Domains

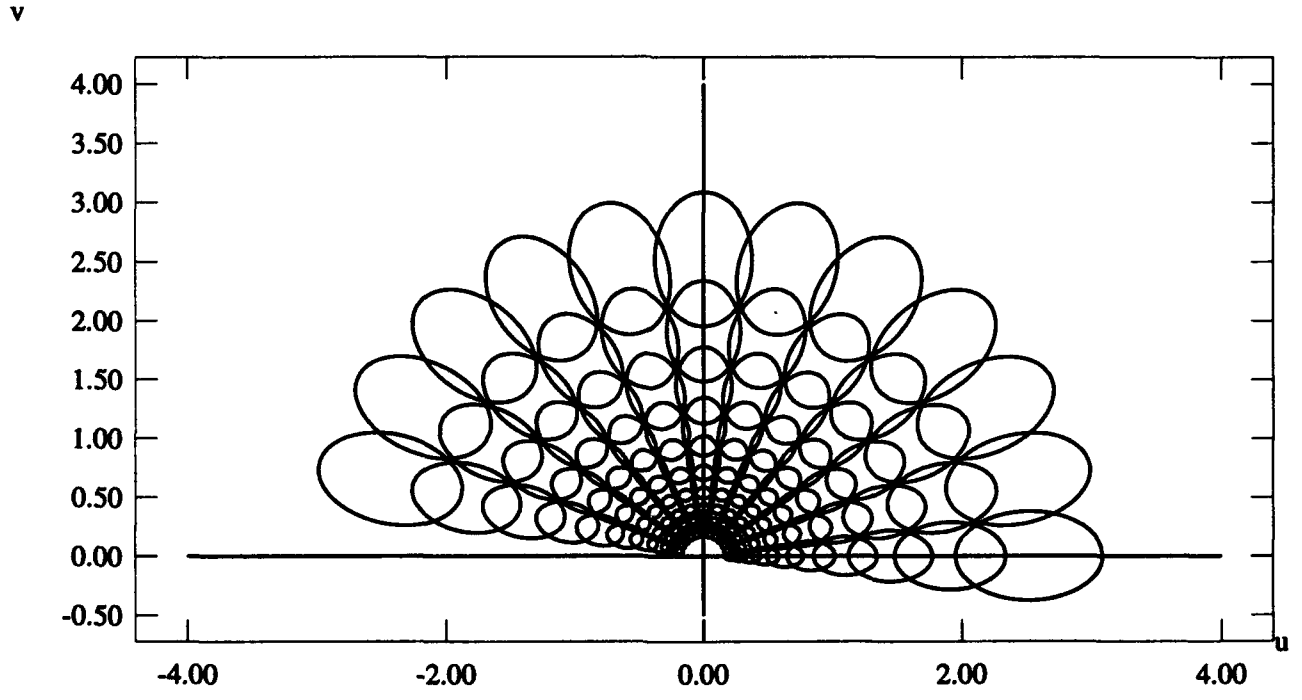


Figure 2: The Set of 2D Variable Window Gabor Filters

based more on very local information. On the contrary, if the image contains strong low frequency components, the final result will be based upon less local information. We no longer need to tune window size every time when processing new images.

4 Stability of Filtering

Though we obtain an amplitude value and a phase value at every pixel location by convolving the image with one filter, both of them may be so severely contaminated by either windowing or noise that it is no longer valid to use them as approximations of the *real* amplitude and phase values. Therefore, any algorithm using them without discretion will potentially result in substantial errors. The goal of this section is to provide a way to quantify such a bias from real values and, subsequently, reduce errors caused by those contaminations.

4.1 Window Contamination

The origin of the window contamination is the convolution in the spatial frequency domain caused by window multiplication in the spatial domain. In other words, we are extracting amplitude and phase values not just at a single frequency, but at a weighted sum of a band of frequencies. The behavior of the sum may or may not be similar to that of a single frequency.

For simplicity, let us first consider the 1D case. For the sum of different frequency components to behave like a single frequency component, we ask the *local* behavior of the sum to satisfy following criteria,

1. The phase should change linearly w.r.t. the position, i.e. the derivative of the phase w.r.t. the pixel position should be the frequency.

2. The phase should be stable w.r.t. the tuning frequency, i.e. the phase should be constant when the tuning frequency is shifted slightly.

3. The amplitude should be stable w.r.t. the position, i.e. the derivative of amplitude w.r.t. the pixel position should be zero.

4. The amplitude should be stable w.r.t. the tuning frequency, i.e. the amplitude should be constant when the tuning frequency is shifted slightly.

Apparently, a complex sinusoid satisfies all the criteria perfectly[2]. Therefore, for the sum to approximate a single sinusoid locally, it has to approximately satisfy those criteria locally. On the other hand, if the sum does satisfy all the constraints, then in the spatial domain, the bandpassed signal can be well approximated by a sinusoid locally. Consequently, we regard locally unstable signals resulting bandpass filterings as those which can not be well approximated by a single sinusoid.

[2]For criterion 4, the amplitude is constant only when the tuned frequency is approximately equal to the frequency of the sinusoid.

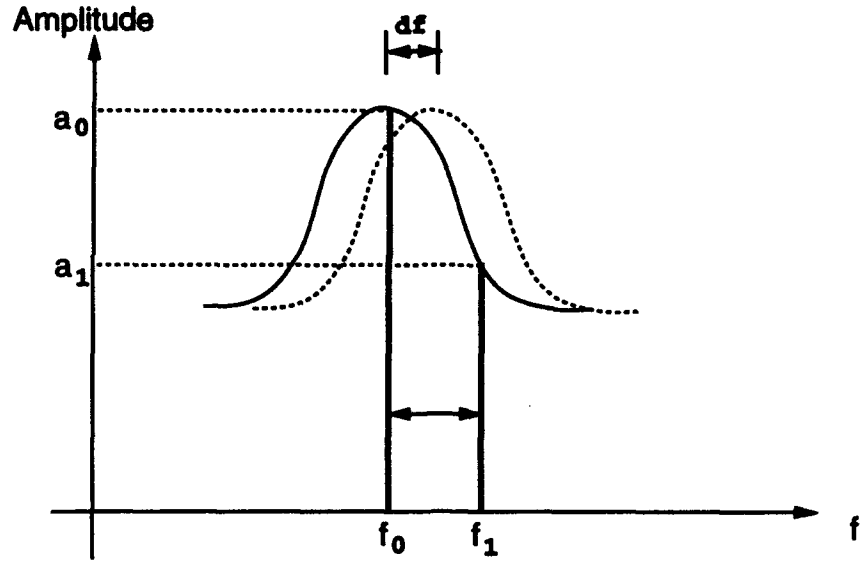


Figure 3: Stability Analysis

Then the task left is to find one or more constraints, which are easy to compute, to indicate to what extent those criteria are satisfied. Let us further simply the problem by assuming the signal is a sum of two independent sinusoids of frequency f_0 and f_1, and weighted amplitude a_0 and a_1. Then the sum is:

$$ae^{j\phi} = a_0 e^{j\phi_0} + a_1 e^{j\phi_1} \tag{5}$$

Let $\frac{a_1}{a_0} = \lambda$, $\Delta\phi = \phi_1 - \phi_0$ and $\Delta f = f_1 - f_0$, after some manipulations, the criteria 1 and 3 can be expressed as:

$$T_1 = \left|\frac{\partial\phi}{\partial x} - f_0\right| = \left|\frac{\lambda\Delta f(\lambda + \cos\Delta\phi)}{\lambda^2 + 2\lambda\cos\Delta\phi + 1}\right| \tag{6}$$

$$T_3 = \left|\frac{1}{a}\frac{\partial a}{\partial x}\right| = \left|\frac{\lambda\Delta f \sin\Delta\phi}{\lambda^2 + 2\lambda\cos\Delta\phi + 1}\right| \tag{7}$$

When the tuning frequency is shifted as in Figure 3, it will affect the amplitude ratio of two components,

$$\frac{d\lambda}{df} = \frac{\frac{a_1 e^{-(\Delta f - df)^2\sigma^2/2}}{e^{-(\Delta f)^2\sigma^2/2}} - \frac{a_1}{a_0}}{df} = \frac{\lambda(e^{(\Delta f)(df)\sigma^2} - 1)}{df} = \lambda\Delta f\sigma^2 \tag{8}$$

where σ is the spatial extent of the filter.

Using Equation 8 and $d(\log f) = \frac{df}{f}$, we can express the criteria 2 and 4 as:

$$T_2 = \left|\frac{\partial\phi}{\partial f}f\right| = \left|f\frac{\partial\phi}{\partial\lambda}\frac{d\lambda}{df}\right|$$

$$= \left| \frac{\sigma^2 f \Delta f \lambda \sin \Delta\phi}{\lambda^2 + 2\lambda \cos \Delta\phi + 1} \right| \tag{9}$$

$$T_4 = \left| \frac{f}{a} \frac{\partial a}{\partial f} \right| = \left| \frac{f}{a} \frac{\partial a}{\partial \lambda} \frac{d\lambda}{df} \right|$$

$$= \left| \frac{\sigma^2 f \Delta f \lambda (\lambda + cos\Delta\phi)}{\lambda^2 + 2\lambda \cos \Delta\phi + 1} \right| \tag{10}$$

Summing those criteria together, we obtained an overall criterion for the stability of the signal which is composed of two independent sinusoids:

$$T^2 = T_1^2 + T_2^2 + T_3^2 + T_4^2 = (1 + \sigma^2 f)^2 (T_1^2 + T_3^2). \tag{11}$$

Apparently, if $T^2 \approx 0$, the four criteria listed previously are approximated satisfied. In other words, the local behavior of the sum is similar to that of a single sinusoid when $T^2 \approx 0$.

The rightmost expression in Eq. 11 provides another benefit for the computation of T, i.e. we don't actually need to compute T_2 or T_4, which usually requires high density frequency sampling because they are derivatives with respect to frequency. In practice, the computations of T_1 and T_3 are straightforward. Note the difference between the criterion proposed here and those in [5]. The criterion proposed here is more general in that only when Eq. 3 is satisfied, they are equivalent, and the criterion in Eq. 11 is for both amplitude and phase information.

Figure 4 illustrates the stability criterion when applied to the sum of a fixed frequency sinusoid and a chirp signal, i.e. the frequency increases linearly with respect to pixel locations, with the same magnitude ($\lambda = 1$). The upper two graphs show Δf and $\Delta\phi$ in Eq. 11, the lower left graph is the signal itself in spatial domain, and the lower right graph shows the computed stability criterion T. Obviously, the spikes in the stability criterion are caused by asynchronous sinusoids, i.e. those sinusoids are concealing each other because the phase difference is approaching $\pm\pi$, and the gradual increase in the stability criterion is caused by the increasing frequency difference.

As proved in Appendix A, Eq. 11 is valid for an arbitrary signal. Generalizing Eq. 11 to any signal, and taking the spatial extent into consideration, we obtain the criterion function T' as in Eq. 12:

$$T' = \frac{T}{k_1} \approx \sqrt{\left(\frac{\partial \phi}{\partial x} - f_0 \right)^2 + \left(\frac{1}{a} \frac{\partial a}{\partial x} \right)^2} \sigma \tag{12}$$

where k_1 is the constant in Eq. 3, and f_0 is the tuning frequency.

Elimination of unstable information can be done by simple thresholding, i.e. if at any location and any frequency band, T' exceeds a certain threshold, then the amplitude and phase information in that frequency band is regarded as unstable. Notice that the threshold should be a constant with respect to different frequency bands.

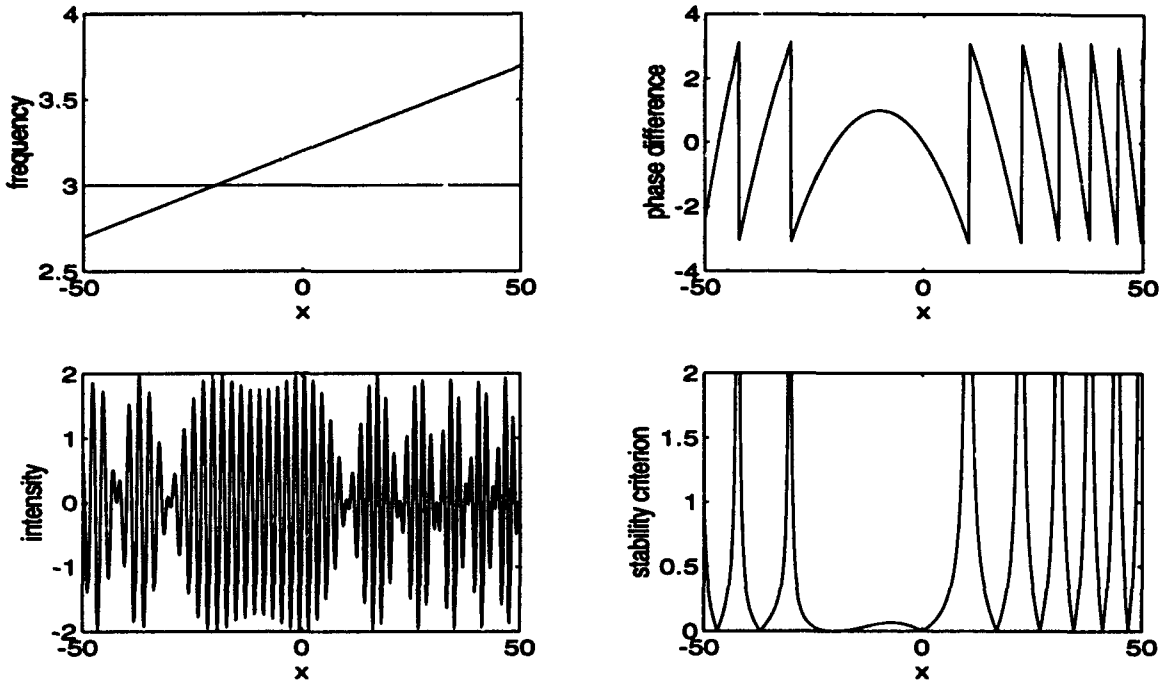


Figure 4: Stability Criterion

In case of 2D Gabor filters proposed previously, let us assume the X axis is coincident with the radial orientation $\mathbf{u}$ of a filter, the four criteria can be expressed as:

$$\mathbf{T}_1 = \nabla_r \phi - \mathbf{u} = \frac{\lambda(\lambda + \cos\Delta\phi)}{\lambda^2 + 2\lambda\cos\Delta\phi + 1} \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix}, \tag{13}$$

$$\mathbf{T}_2 = u\nabla_u \phi = \frac{\lambda u\sigma^2 \sin\Delta\phi}{\lambda^2 + 2\lambda\cos\Delta\phi + 1} \begin{pmatrix} \Delta u \\ \frac{\Delta v}{k_2^2} \end{pmatrix}, \tag{14}$$

$$\mathbf{T}_3 = \frac{1}{a}\nabla_r a = \frac{\lambda \sin\Delta\phi}{\lambda^2 + 2\lambda\cos\Delta\phi + 1} \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix}, \tag{15}$$

$$\mathbf{T}_4 = \frac{u}{a}\nabla_u a = \frac{\lambda u\sigma^2(\lambda + \cos\Delta\phi)}{\lambda^2 + 2\lambda\cos\Delta\phi + 1} \begin{pmatrix} \Delta u \\ \frac{\Delta v}{k_2^2} \end{pmatrix}, \tag{16}$$

where ∇_r is the gradient operator in spatial domain, ∇_u is the gradient operator in spatial frequency domain, $\mathbf{u}$ is the peak tuning frequency, Δu and Δv are the frequency difference in X and Y directions, and k_2 is the aspect ratio of the filter. Similarly the overall stability criterion can be expressed as:

$$\|T'\|^2 = \sigma^2(\|\mathbf{S}\mathbf{T}_1\|^2 + \|\mathbf{S}\mathbf{T}_3\|^2), \tag{17}$$

Where the $\mathbf{S}$ is the aspect scaling matrix,

$$\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{k_2^2} \end{pmatrix}$$

4.2 Noise Contamination

Due to different window sizes of filters, the effects of noise on the amplitude and phase values are also different. Assuming additive white noise, the noise component included in filter output can be expressed as:

$$N(f_0) = \int_{-\infty}^{+\infty} e^{-(f-f_0)^2\sigma^2/2} n e^{j\phi_n} df = \frac{\sqrt{2\pi}n}{\sigma} \int_{-\infty}^{+\infty} g(f - f_0; \frac{1}{\sigma}) e^{j\phi_n} df, \qquad (18)$$

where $g(f - f_0; \frac{1}{\sigma})$ is a Gaussian function with center at f_0 and extent $\frac{1}{\sigma}$, n is the magnitude of noise at every frequency, ϕ_n is the random phase of noise.

Because we assumed white noise, the expected value of the rightmost integral term in Eq. 18 is assumed to be a complex number with constant magnitude and random phase. Therefore, using Eq. 2, we obtain the relative noise level in every frequency band.

$$\|N(f_0)\| = k_3 f_0, \qquad (19)$$

or in the 2D case,

$$\|N(\mathbf{f_0})\| = k_3 \|\mathbf{f_0}\|^2, \qquad (20)$$

where k_3 is a constant.

5 Applications

The general framework developed in this paper is applicable to any vision task which makes use of either amplitude or phase information. We will explain two specific applications, focus quality measurement and 2D correspondence. The reason we choose these two applications is as we will see, focus quality measurement makes use of amplitude information only, and 2D correspondence makes use of phase information only.

5.1 Focus Quality Measurement

As explained in [22, 17], the key problem in focus quality measurement can be stated as, given two images which are blurred to different extent, how to measure *locally* the difference of blurring at each pixel location. Modeling the blurring as a convolution with a Gaussian, we have[17],

$$\sigma' = \frac{1}{f}\sqrt{|\ln \|I_0(f)\|^2 - \ln \|I_1(f)\|^2|}, \qquad (21)$$

where σ' represents the blurring difference, and $\|I_0(f)\|$ and $\|I_0(f)\|$ are amplitude values of two images at frequency f.

We can apply the set of 2D variable window Gabor filters to extract amplitude information at each frequency band, eliminating unstable amplitude information by

thresholding in Eq. 17, then fit all other stable amplitude values and their included noise level (Eq. 20) into the following linear (w.r.t. σ'^2) equation[21].

$$\sigma'^2\|\mathbf{f}\|^2 = \left|\ln\|I_0(f)\|^2 - \ln\|I_1(f)\|^2\right| + c, \tag{22}$$

where the constant c is used to compensate illumination difference between two images, and instead of the peak tuning frequency, $\|\mathbf{f}\|^2$ can better approximated by the average of two instantaneous frequencies $\|\frac{1}{2}(\nabla_r\phi_0 + \nabla_r\phi_1)\|^2$. Ordinary χ^2 estimation can be applied to the fitting with the uncertainty of the right side being expressed as [21],

$$\frac{\|N(\mathbf{f})\|}{\|I_0(\mathbf{f})\|} + \frac{\|N(\mathbf{f})\|}{\|I_1(\mathbf{f})\|}. \tag{23}$$

The iterative estimation in [22] can be also efficiently implemented as convolutions in the spatial frequency domain.

$$I(x) \otimes g(x;\sigma') \otimes f(x) = (I(x) \otimes f(x)) \otimes g(x;\sigma') \tag{24}$$

where $f(x)$ is a bandpass filter.

5.2 2D Correspondence

The key problem in 2D correspondence is to find the spatial shift between two images at each pixel location. Because a spatial shift is equivalent to a phase shift in frequency domain, we can infer the spatial shift from phase difference [18, 7, 15].

An obvious approach similar to that of focus quality measurement is to apply the set of 2D variable window Gabor filters to extract phase information at each frequency band, eliminate unstable phase information by thresholding in Eq. 17, then minimize the following to find disparity $\mathbf{r}$.

$$\sum_{\phi_0\phi_1 \text{ stable}} \frac{((\phi_1 - \phi_0 - \mathbf{f}\cdot\mathbf{r}) \bmod 2\pi)^2}{\left(\frac{\|N(\mathbf{f})\|}{\|I_0(\mathbf{f})\|} + \frac{\|N(\mathbf{f})\|}{\|I_1(\mathbf{f})\|}\right)^2}, \tag{25}$$

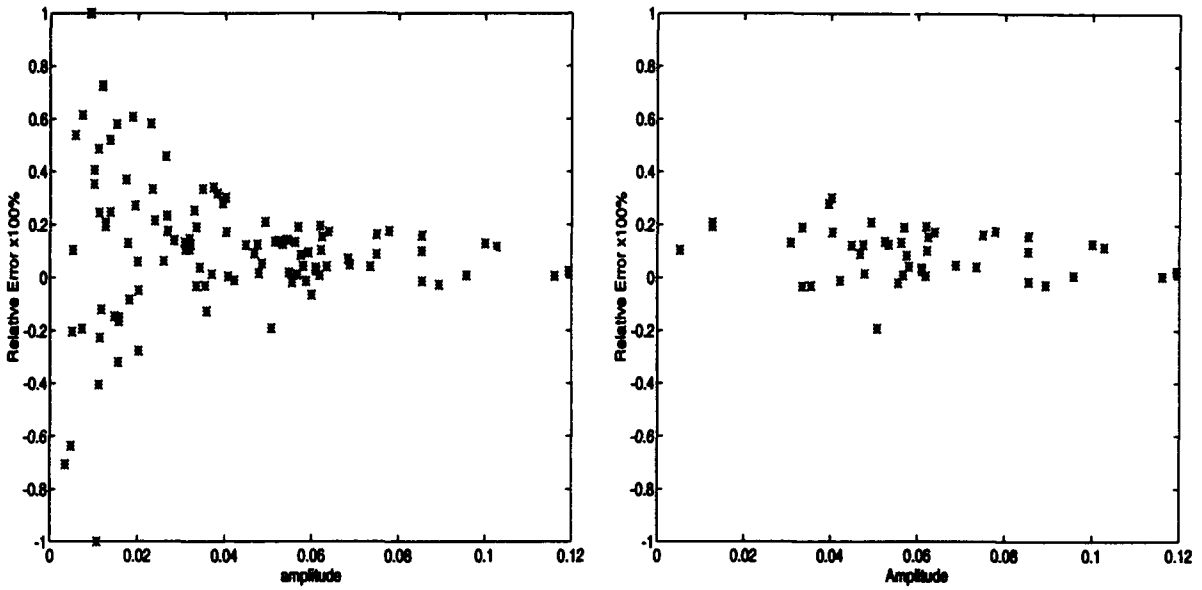
where $N(\mathbf{f})$ is the noise component in Eq. 20.

The iterative estimation in [7] can be done by simply shifting one image locally according to previously estimated disparities.

6 Experimental Results

6.1 Elimination of Unstable Information

First we artificially convolved an image with a Gaussian function $g(x;\sigma = 1.0)$, then convolved the set of filters with original and blurred images, and analyzed the relations between amplitudes of various frequencies at an arbitrarily chosen pixel location. The left half of Figure 5 shows the relative error of estimating σ (Eq. 22) from a single



Before Elimination After Elimination

Figure 5: Elimination of Unstable Amplitude Information

frequency band without eliminating unstable amplitude information[3], and the right half shows the error after thresholding by $T' = 1.25$ in Eq. 17.

To illustrate the effectiveness of eliminating unstable phase information, we spatially shifted a image by $\mathbf{r}$, and then the phase difference φ represents the error of phase with respect to ideal sinusoids.

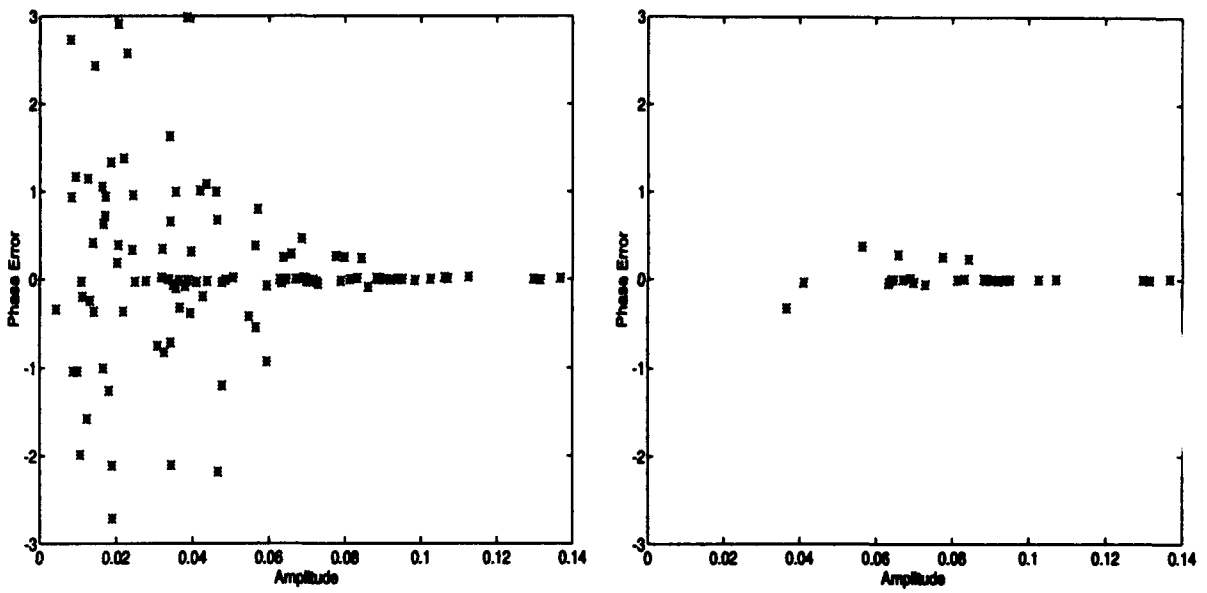
$$\varphi = \left(\mathbf{r} \cdot \frac{\nabla\phi_1 + \nabla\phi_2}{2} - (\phi_1 - \phi_2) \right) \bmod 2\pi, \tag{26}$$

where ϕ_1 and ϕ_2 are the two phase values resulting from convolving a filter with the original and shifted images. Ideally, φ should be zero at any frequency band. At an arbitrarily chosen pixel location, the left half of Figure 6 shows φ of every frequency band without eliminating unstable phase information, and the right half shows φ after thresholding by $T' = 1.25$ in Eq. 17.

As illustrated in Figure 5 and Figure 6, the identification of unstable information is indeed very accurate. Depending on a specific application, the threshold can be changed to satisfy looser or tighter requirement for the output of a bandpass filter to resemble a sinusoid.

As illustrated in Figure 5 and Figure 6, there also exists a strong correlation between instability and amplitude values. This suggests an amplitude thresholding scheme ([21]), which assumes unstable amplitude and phase information is caused by low amplitude values. Even though this amplitude thresholding scheme can indeed works well in some cases, when we compared with the stability thresholding, we found

[3]In this case c in Eq. 22 is zero.



Before Elimination After Elimination

Figure 6: Elimination of Unstable Phase Information

that it usually requires images have strong high frequency component because otherwise weak high frequency components are always below the threshold even though they *are* stable with respect to the stability criterion. On the other hand, if we lower the amplitude threshold, a large number of unstable information will get through. In the experiments described below, the amplitude threshold is set to 0.06 (refer to Figure 5 and Figure 6), assuming images are normalized so that their DC components have amplitude value of 1.0.

Figure 7 shows two images of two textured surfaces at difference depth. The images are taken under different lens apertures, which make the difference of focus quality a function of depth [21]. We applied the set of filters to the two images, then eliminated unstable amplitude information by thresholding the stability criterion, and fitted a line against Eq. 22 to obtain σ' at each pixel location. Figure 8 shows the difference of focus quality in the rectangular region in Figure 7. As we can see, the two surfaces are separated obviously, and the depth discontinuity is well located. Figure 9 shows the computed focus quality difference by amplitude thresholding. The area with strong high frequency components are recovered correctly, while in other areas, the results show instability.

Figure 10 shows two images of a rotating ball. We applied the techniques of filtering and stability analysis to the two images, and use Eq. 25 to find 2D disparities. Figure 11 shows the 2D correspondence between the two images. Intrinsic to the algorithm itself, subpixel accuracy is achieved without any interpolation. Given sparse features as those on the ball, the algorithm can automatically avoid unstable information obtained in featureless areas. Figure 12 shows the 2D correspondence using the amplitude threshold. Around the boundary of the ball, where the contrast is lower,

Figure 7: Two Images Taken Under different Lens Aperture

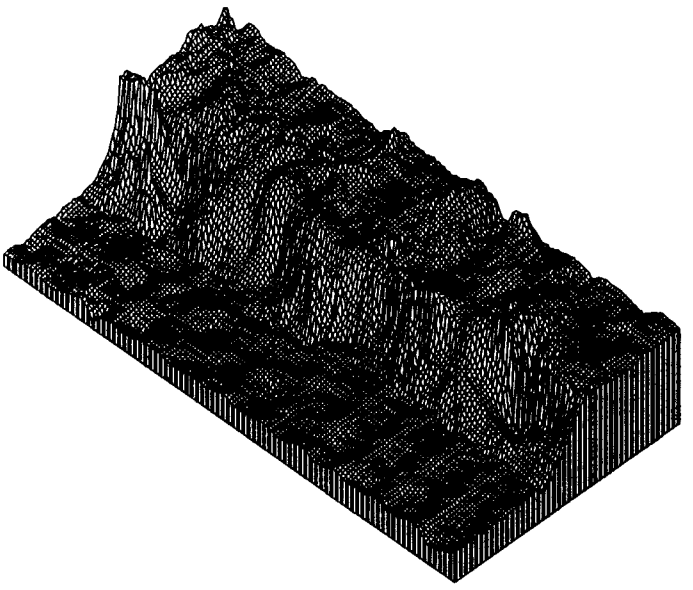


Figure 8: Difference of Focus Quality Using Stability Threshold

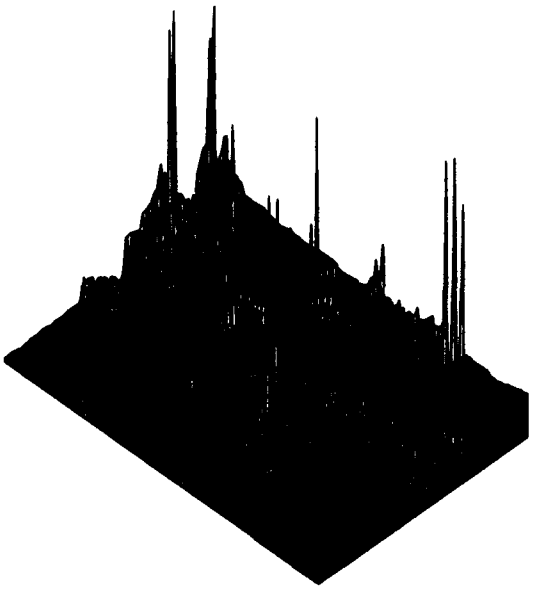


Figure 9: Difference of Focus Quality Using Amplitude Threshold

the algorithm generate unstable results.

6.2 Fixed Window vs. Variable Window

The selection of window size for any window operation is a compromise between resolution and stability. While large windows reduce effects of noise, and make results more stable, the resolution is reduced at the same time. On the other hand, small windows have the advantage of preserving locality, i.e. high resolution, they are potentially unstable. Therefore a proper window size has to be *ad hoc* to a specific problem. An implicit rule of choosing a window size is to select a window which is as small as possible while the error caused by noise and windowing is still within an acceptable range.

The alternative way of selecting a window size every time when processing new images, is to adopt the variable window scheme proposed in this paper. Because window sizes are proportional to wavelengths, this variable window scheme uses large windows when frequencies are low, small windows when frequencies are high.

Figure 13 and Figure 14 show an example of input images with different contents. The images in Figure 13 contain a considerable amount of high frequency information, while the images in Figure 14 contain only low frequency information. The black lines in the images are the locations where focus quality differences are measured. Focus quality measurements are done using the fixed window scheme [22] with a large window ($\sigma = 20.0$) and a small window ($\sigma = 10.0$), and the variable window scheme described in previous sections.

Figure 15 shows focus quality difference of a single scan line between images in Figure 13. The large window scheme resulted in a large slope around the depth

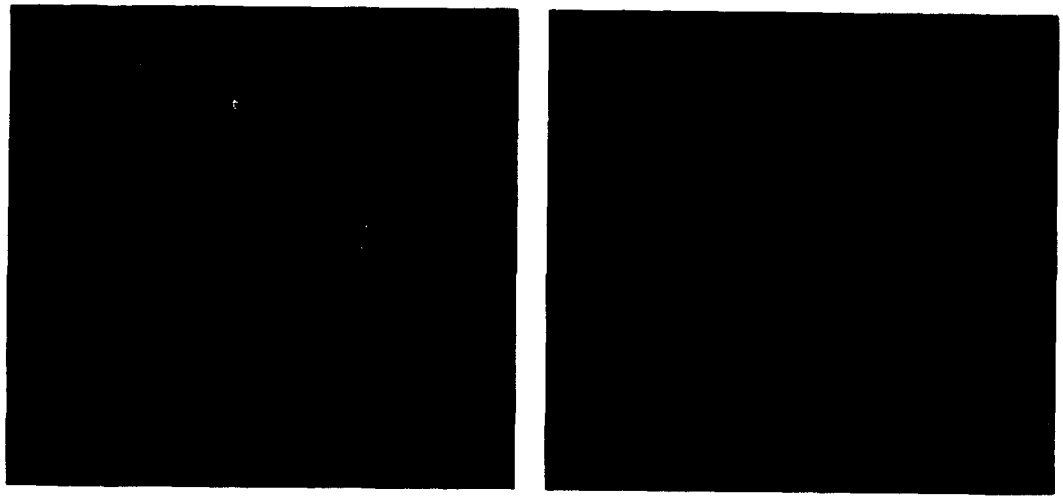


Figure 10: Two Images of a Ball

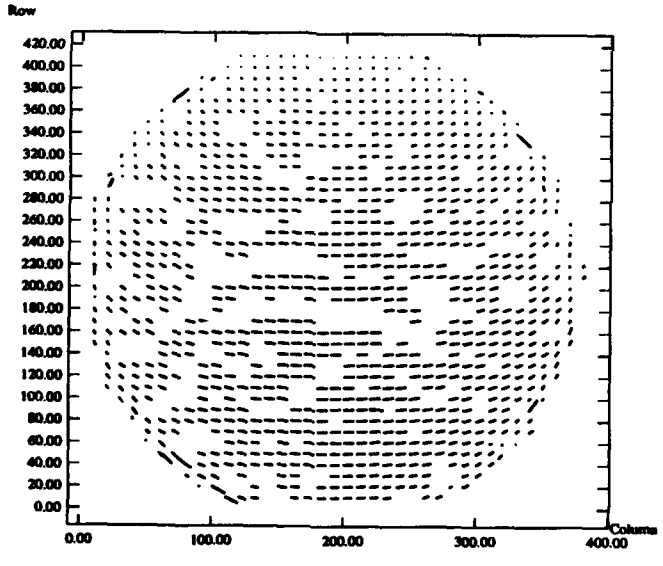


Figure 11: 2D Correspondence Using Stability Threshold

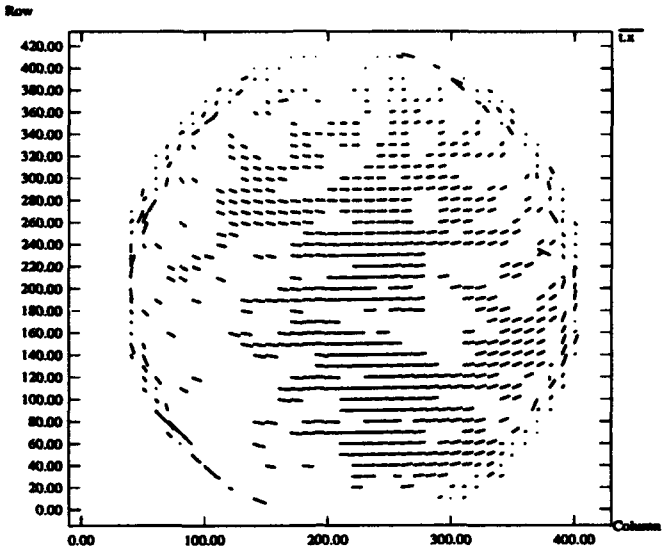


Figure 12: 2D Correspondence Using Amplitude Threshold

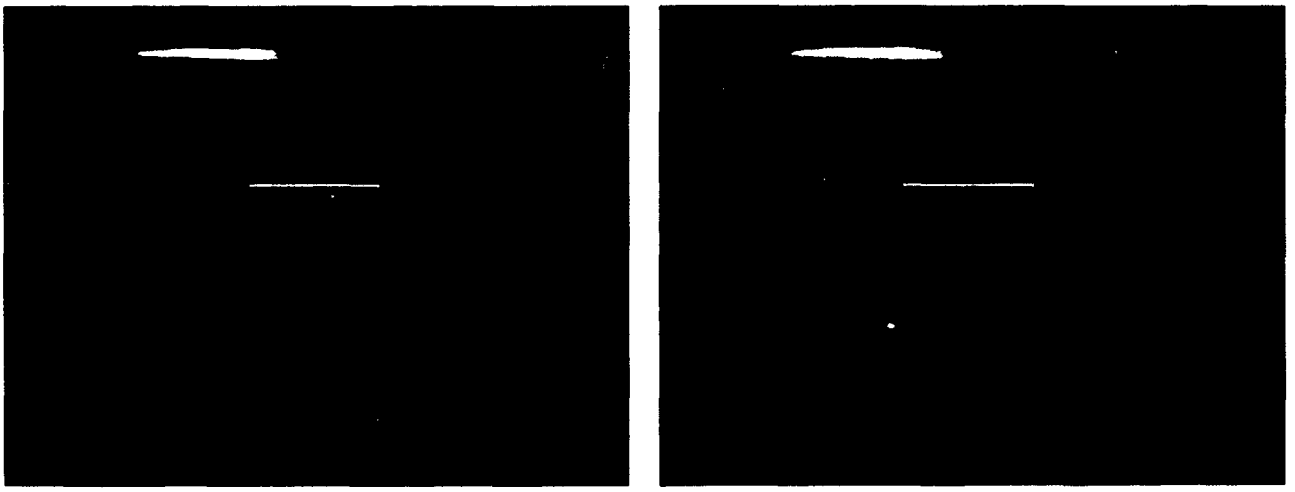


Figure 13: Nearly Focused Image Pair

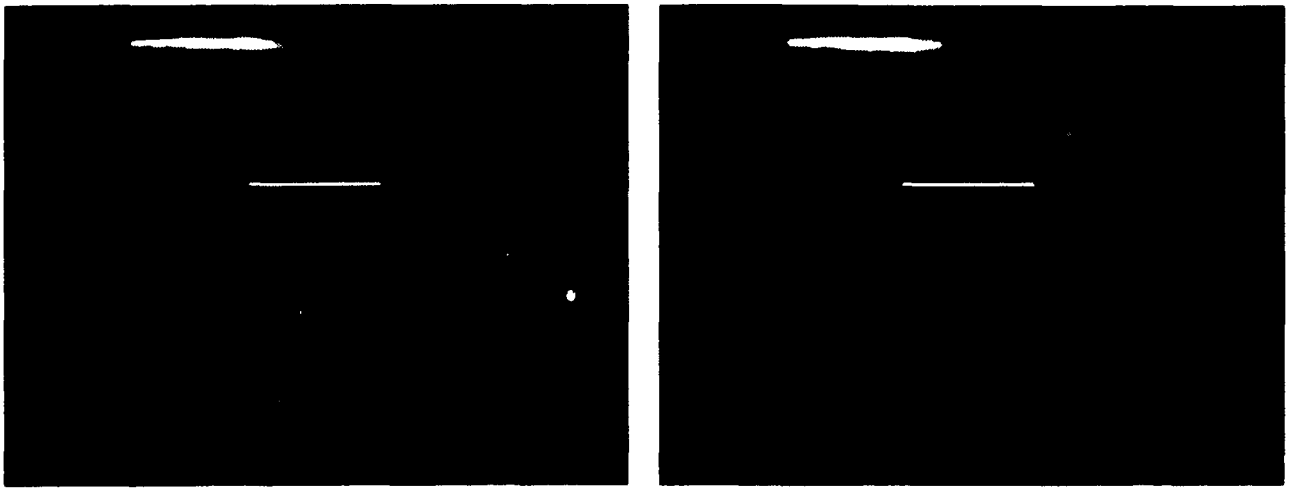


Figure 14: Defocused Image Pair

	Nearly Focused Image Pair	Defocused Image Pair
Small Fixed Window	0.1218	0.2708
Large Fixed Window	0.1677	0.1947
Variable Window	0.1150	0.1664

Table 1: RMS Errors of Different Window Schemes

discontinuity. The small window scheme displayed more noisy results before and after the discontinuity even though it located the discontinuity more precisely. Figure 16 showed the results of focus quality difference between images in Figure 14. This time, the small window scheme generated unstable results, and the large window scheme produced a blurred, but stable results. In both Figure 15 and Figure 16, the variable scheme resulted in a meaningful compromise between resolution and precision. Table 1 shows root mean square errors of all cases. The variable window scheme performs significantly better than the large window scheme in the nearly focused image pair, and the small window scheme in the defocused image pair.

7 Conclusion

This paper provide a general framework of visual computing in spatial frequency domain. We addressed two main problems in spatial frequency analysis, one is the window size problem, the other is the stability problem associated with windowing. We compared the variable window scheme with the fixed window scheme, and introduced a more general stability criterion. And we showed the effectiveness of these two tools in applications of focus quality measurement and 2D correspondence.

Acknowledgement

The authors would like to thank Mark Maimone for helpful discussions. This re-

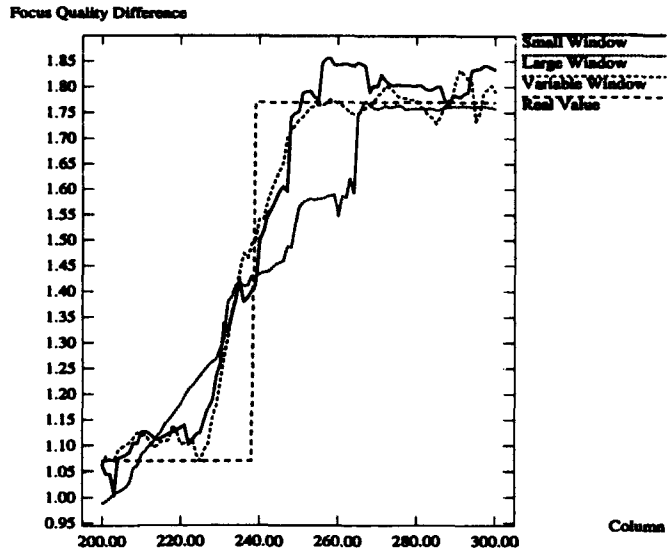


Figure 15: Focus Quality Difference of the Nearly Focused Images

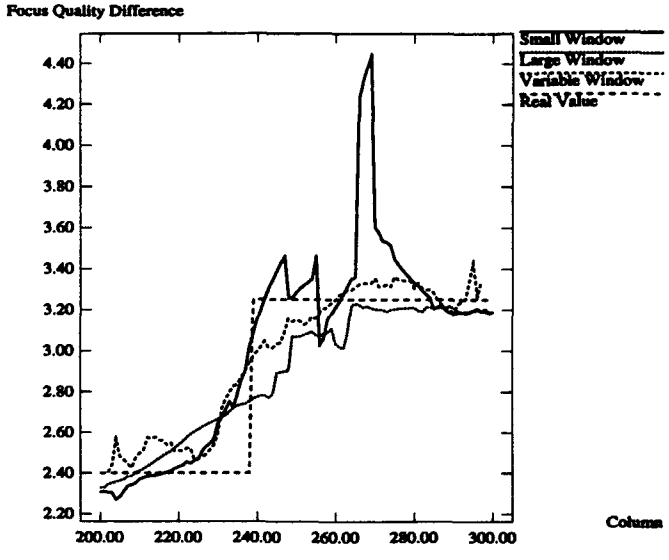


Figure 16: Focus Quality Difference of the Defocused Images

search was sponsored by the Avionics Lab, Wright Research and Development Center, Aeronautical Systems Division (AFSC), U. S. Air Force, Wright-Patterson AFB, OH 45433-6543 under Contract F33615-90-C-1465, ARPA Order No. 7597. The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of DARPA or the U. S. Government.

A Stability Criteria for An Arbitrary Signal

Let us assume an arbitrary signal $i(x)$, and its Fourier transform

$$I(f) = \int_{-\infty}^{+\infty} i(x) e^{-jfx} dx. \tag{27}$$

Its Gabor transform in any location x_0 and any peak frequency f_0 can be represented as:

$$\begin{aligned} G(x_0, f_0) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} i(x - x_0) e^{-\frac{x^2}{2\sigma^2}} e^{-jf_0 x} dx \\ &= \int_{-\infty}^{+\infty} I(f) e^{-jfx_0} e^{-(f_0 - f)^2\sigma^2/2} df \\ &= a(x_0, f_0) e^{j\phi(x_0, f_0)} \\ &= R_0(x_0, f_0) + jI_0(x_0, f_0) \end{aligned}$$

Then, we have

$$\frac{\partial \phi}{\partial x_0} = \frac{R_0 \frac{\partial I_0}{\partial x_0} - I_0 \frac{\partial R_0}{\partial x_0}}{R_0^2 + I_0^2} \tag{28}$$

$$\frac{1}{a}\frac{\partial a}{\partial x_0} = \frac{R_0 \frac{\partial R_0}{\partial x_0} + I_0 \frac{\partial I_0}{\partial x_0}}{R_0^2 + I_0^2} \tag{29}$$

Similarly,

$$f_0 \frac{\partial \phi}{\partial f_0} = f_0 \frac{R_0 \frac{\partial I_0}{\partial f_0} - I_0 \frac{\partial R_0}{\partial f_0}}{R_0^2 + I_0^2} \tag{30}$$

$$\frac{f_0}{a}\frac{\partial a}{\partial f_0} = f_0 \frac{R_0 \frac{\partial R_0}{\partial f_0} + I_0 \frac{\partial I_0}{\partial f_0}}{R_0^2 + I_0^2} \tag{31}$$

Let's define $G_1(x_0, f_0)$:

$$G_1(x_0, f_0) = \frac{j}{\sqrt{2\pi}\sigma^3} \int_{-\infty}^{+\infty} i(x - x_0) x e^{-\frac{x^2}{2\sigma^2}} e^{-jfx} dx \tag{32}$$

$$= \int_{-\infty}^{+\infty} I(f) e^{-jfx_0} (f_0 - f) e^{-(f_0 - f)^2\sigma^2/2} df \tag{33}$$

$$= R_1 + jI_1,$$

where the equivalence between Eq. 33 and Eq. 32 can be easily verified by using the following property of Fourier transform:

$$\mathcal{F}[gh] = \mathcal{F}[g] \otimes \mathcal{F}[h], \tag{34}$$

where $\mathcal{F}$ denotes the Fourier transform operator.

Now, we have,

$$\begin{aligned}
\frac{\partial G}{\partial x_0} &= \frac{\partial R_0}{\partial x_0} + j\frac{\partial I_0}{\partial x_0} \\
&= -\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} i'(x-x_0)e^{-\frac{x^2}{2\sigma^2}}e^{-ijf_0x}dx && (35) \\
&= -\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} i(x-x_0)(\frac{x}{\sigma^2}+jf_0)e^{-\frac{x^2}{2\sigma^2}}e^{-jf_0x}dx && (36) \\
&= j(G_1(x_0,f_0) - f_0G(x_0,f_0)), && (37)
\end{aligned}$$

where, from step 35 to step 36, we used the partial integration method:

$$\int u'(x)v(x)dx = u(x)v(x) - \int u(x)v'(x)dx$$

Using the same technique, we have,

$$\begin{aligned}
\frac{\partial G}{\partial f_0} &= \frac{\partial R_0}{\partial f_0} + j\frac{\partial I_0}{\partial f_0} \\
&= -\int_{-\infty}^{+\infty} I(f)e^{-jfx_0}\sigma^2(f_0-f)e^{-(f_0-f)^2\sigma^2/2}df \\
&= -\sigma^2 G_1(x_0,f_0) && (38)
\end{aligned}$$

Replacing Eq. 37 and Eq. 38 into Eq. 28 29 30 31, we obtain,

$$T_1 = \frac{\partial\phi}{\partial x_0} - f_0 = \frac{R_0R_1 + I_0I_1}{R_0^2 + I_0^2} \tag{39}$$

$$T_2 = f_0\frac{\partial\phi}{\partial f_0} = \frac{\sigma^2 f_0(I_0R_1 - R_0I_1)}{R_0^2 + I_0^2} \tag{40}$$

$$T_3 = \frac{1}{a}\frac{\partial a}{\partial x} = \frac{I_0R_1 - R_0I_1}{R_0^2 + I_0^2} \tag{41}$$

$$T_4 = \frac{f_0}{a}\frac{\partial a}{\partial f_0} = -\frac{\sigma^2 f_0(R_0R_1 + I_0I_1)}{R_0^2 + I_0^2} \tag{42}$$

Therefore, we have,

$$T_1^2 + T_3^2 = \frac{1}{(\sigma^2 f_0)^2}(T_2^2 + T_4^2) = \frac{R_1^2 + I_1^2}{R_0^2 + I_0^2} \tag{43}$$

References

[1] Edward H. Adelson and James R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of Optical Society of America, A*, pages 284–299, 1984.

[2] V. Michael Bove, Jr. Discrete fourier transform based depth-from-focus. In *Proceedings OSA Topical Meeting on Image Understanding and Machine Vision*, 1989.

[3] Alan C. Bovik, Marianna Clark, and Wilson S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Trans. Patt. Recog. Machine Intell.*, 12(1):55–73, January 1990.

[4] John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of Optical American Society A*, 2(7), 1985.

[5] David Fleet and Allan Jepson. Stability of phase information. In *Proc. of IEEE Workshop on Visual Motion*, pages 52–60, Princeton, New Jersey, October 1991.

[6] David J. Fleet and Allan D. Jepson. Computation of normal velocity from local phase information. In *Proc. Comput. Vision Patt. Recog.*, pages 379–386, 1989.

[7] David J. Fleet, Allan D. Jepson, and Michael Jenkin. Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2):198–210, 1991.

[8] D. Gabor. Theory of communication. *Journal of the IEE*, 93:429–457, 1946.

[9] David J. Heeger. Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, pages 279–302, January 1988.

[10] Anil K. Jain and Farshid Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[11] David G. Jones and Jitendra Malik. A computational framework for determining stereo correspondence from a set of linear spatial filters. In *European Conference on Computer Vision*, pages 395–410, 1992.

[12] John Krumm and Steven A. Shafer. Local spatial frequency analysis for computer vision. Technical Report CMU-RI-TR-90-11, The Robotics Institute, Carnegie Mellon University, 1990.

[13] John Krumm and Steven A. Shafer. Shape from periodic texture using the spectrogram. In *Proc. Computer Vision Patt. Recog.*, pages 284–289, 1992.

[14] John Krumm and Steven A. Shafer. Segmenting textureed 3D surfaces using the space/frequency representation. Technical Report CMU-RI-TR-93-14, The Robotics Institute, Carnegie Mellon University, 1993.

[15] K. Langley, T. J. Atherton, R. G. Wilson, and M. H. E. Larcombe. Vertical and horizontal disparities from phase. In *European Conference on Computer Vision*, pages 315–325, 1990.

[16] Jitendra Malik and Ruth Rosenholtz. A differential method for computing local shape-from-texture for planar and curved surfaces. In *Proc. Computer Vision Patt. Recog.*, pages 267–273, 1993.

[17] Alex P. Pentland. A new sense for depth of field. *IEEE Transactions on PAMI*, 9(4):523–531, 1987.

[18] T. D. Sanger. Stereo disparity computation using Gabor filters. *Biological Cybernetics*, 59:405–418, 1988.

[19] Murali Subbarao. Parallel depth recovery by changing camera parameters. In *2nd International Conference on Computer Vision*, pages 149–155, 1988.

[20] Juyang Weng. A theory of image matching. In *International Conference on Computer Vision*, pages 200–209, 1990.

[21] Yalin Xiong and Steven Shafer. Depth from focusing and defocusing. Technical Report CMU-RI-TR-93-07, The Robotics Institute, Carnegie Mellon University, 1993.

[22] Yalin Xiong and Steven A. Shafer. Depth from focusing and defocusing. In *Proc. Computer Vision Patt. Recog.*, pages 68–73, 1993.